



Proposal and Data Exploration for PREDICT-wide Modeling & Analytics Project: Seasonality and Other Factors Affecting Viral Shedding in Bats

Project Rationale

Recent examples of emerging viruses with serious public health consequences, including SARS, Ebola, and Zika, have highlighted the need for a better understanding of the wildlife reservoirs of these pathogens and the processes that result in their spillover into humans. An obvious and necessary condition enabling spillover is active infection in the animal species that serve as pathogen reservoirs. Thus, knowledge of spatial and temporal dynamics of natural infection in wildlife reservoirs should allow for more effective, targeted surveillance, strategies to avoid high risk periods and regions, and generally improve our ability to predict and manage viral spillover.

Bats are a particularly important reservoir host taxon given their propensity to carry zoonotic viruses and their involvement in spillover of prominent EIDs. A central question regarding viral dynamics in bat hosts is whether bats are more likely to be infected during certain time periods and, if so, why. For example, hypotheses have been published that link seasonal migration, hibernation, communal synchronous birthing and other factors, but it is unknown how significant these issues are to viral dynamics, or how widespread. Seasonal viral shedding patterns have been found in several bat-virus systems, yet the potential mechanisms driving these patterns are diverse and not well resolved. For example, seasonal breeding activity alone could be linked to infection status and viral loads in bat populations through multiple, complementary processes, including:

- Increased population density and intraspecific contact, allowing for pathogen transmission
- An influx of susceptibles into the population
- Changes in individual reproductive or nutritional status that affect susceptibility

In this proposed work we aim to work with key staff across the PREDICT partnership to develop a generalizable, analytical framework that can parse out the relative influence of these factors on bat viral shedding while controlling for confounding factors. Testing this framework using the large PREDICT dataset (i.e. not limited to single bat host species or a single geographic region, or single year) will allow for broader generalizations about seasonality of viral shedding than has ever previously been possible.

Project Outline

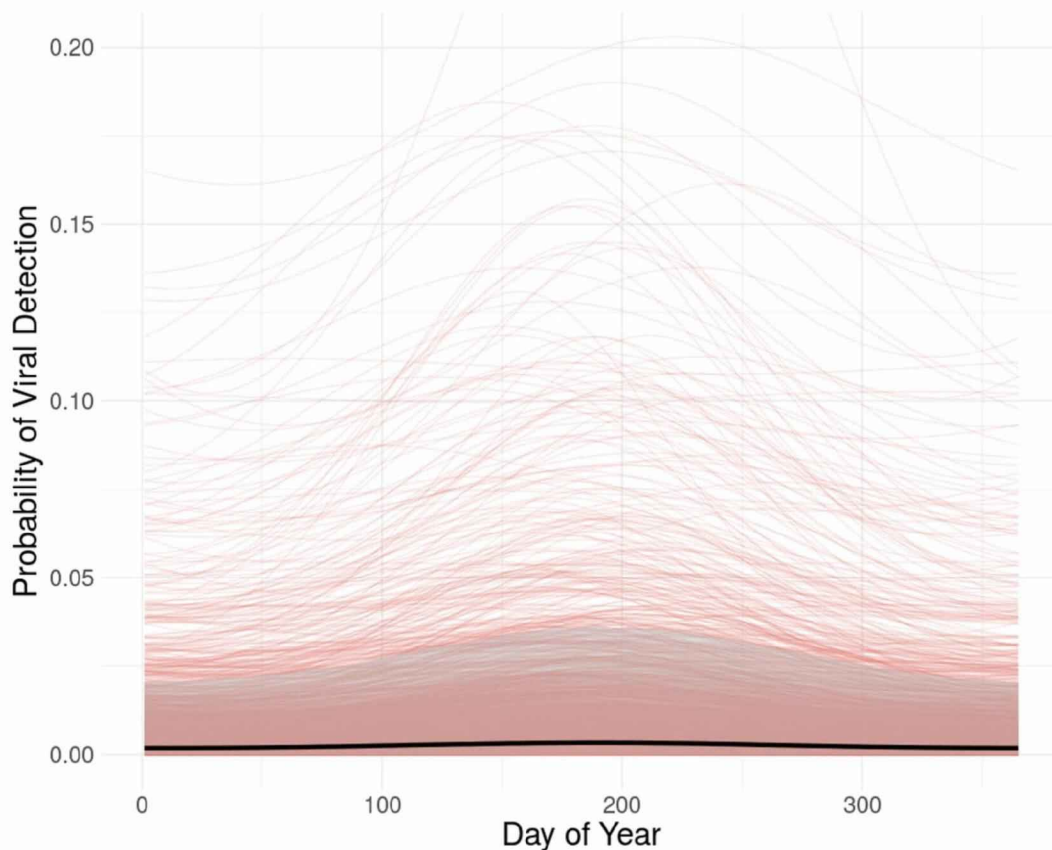
To address these knowledge gaps, we propose an analysis using PREDICT-1 (project-wide, government approved) bat viral detection data collated in EIDITH. PREDICT-1 is a heterogeneous data source, and thus our general strategy will be to use Bayesian modeling implemented within the Stan programming language in order to enable the fullest use of data while accurately representing uncertainty in parameter estimates and predictions

where data is sparse. In addition, we will control for some variability by initially limiting our dataset to conventional PCR tests (excluding real-time PCR, sequencing, and serology results). An outline of our analysis strategy and relevant variables is as follows:

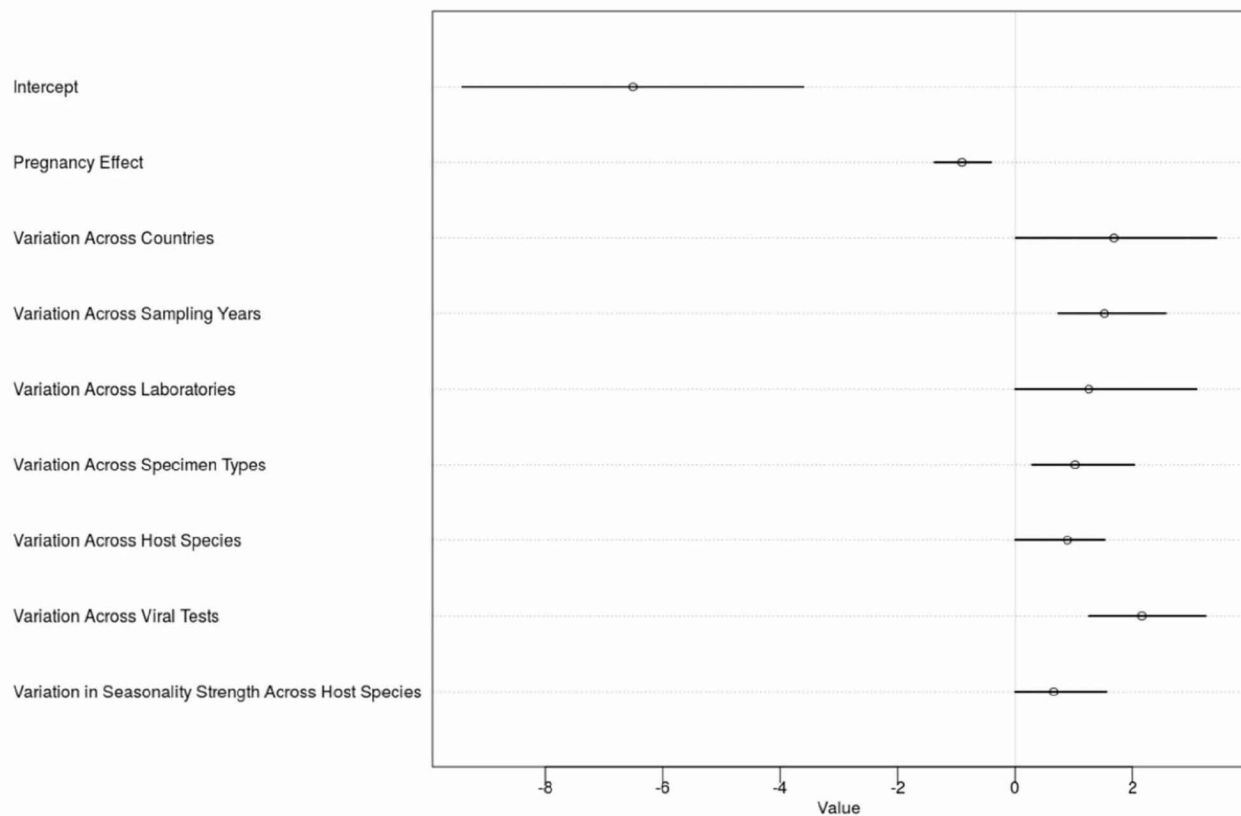
- The binary outcome variable for all models will be viral detection, at the level of the individual PCR test
- All models will allow for seasonal trends in viral shedding, modeled using a sine function
 - By defining the sine function to have yearly periodicity, we can sensibly model seasonal fluctuations across the year. Thus, our modeling will describe smooth changes in viral detection across day of year (0 to 365) rather than relying on coarser, categorical descriptors of seasonality (i.e., wet vs. dry season or month of year). This flexible framework is needed as PREDICT data have been collected from various latitudes where wet/dry seasons vary across the calendar month. It will allow us to interpret these differences and their implications more easily and rapidly.
 - Bayesian modeling techniques allow us to incorporate data across host species while making strong inferences only in cases where there is data support (i.e. “seasonality” for data-poor species will effectively shrink to constant viral detection across the year except in cases where data provide evidence of strong seasonal trends)
- All models will control for variation (by incorporating varying intercepts) across important categorical variables within the EIDITH dataset, including: bat host species from which the specimen was collected, country of specimen collection, year of specimen collection, specimen type, viral family tested for, and the diagnostic laboratory conducting the testing
 - Parameter estimates for these variables can be explicitly interpreted in cases where there is clear biological meaning underlying potential variation (e.g. differences in viral detection across bat host species may represent inherent differences in viral susceptibility) or they can be treated solely as variables that are being controlled for (e.g. differences in viral detection across years may simply represent stochastic variation)
- Models will be constructed to investigate the relative explanatory power of additional predictor variables (main effects) on viral detection, including: rainfall (which drives food availability and potentially social behaviors), lactation status of adult female bats, and pregnancy status of adult female bats
 - While the above models are intended to apply specifically to adult female bats, whose reproductive status is clearly defined, alternative models could be constructed including data on male and/or juvenile bats that incorporate predictors related to adult female reproductive status. These models would address potential effects of reproductive seasonality/social behaviors on viral shedding at the population-level. This is important because reproductive activity in a given species at a given time may serve as a proxy for social interactions and contact that promotes viral sharing and therefore affects both sexes and potentially multiple age classes.
- Analogous models run on subsets of the data (i.e., subset down to a single bat host species or viral family) could be used to validate more general results from global models and strengthen inferences

Although results would be most robust and useful when incorporating as much PREDICT-1 data as possible, our methodology allows for easy iteration and refinement of the exact model structure and underlying data if there are data sensitivities.

Preliminary Results (Using EHA-only Data)



Example Figure 1. Seasonal probability of viral detection in the bat species *Pteropus giganteus*. Estimates of the seasonal trend in viral detection for *P. giganteus* (using only adult female data) were generated from a Bayesian model run using Stan for a total of 4000 iterations. Each red line represents the seasonal trend implied by the parameter estimates of a single model iteration. Thus, there are 4000 seasonal trends plotted, representing the full uncertainty inherent in the modeling process. These seasonal trends are fit using testing data spanning multiple viral families (13 in total), hence our results represent general patterns in viral dynamics. The black line represents the median seasonal trend derived from these estimates. The gray shaded area represents the 95% highest posterior density interval for the seasonal trends. This 95% interval highlights a peak in viral detection at approximately day of year 200. The more data that are used for modeling, the more likely we are to be able to finely resolve seasonal trends across species. Note that viral detection probability ranges from 0.00 - 0.20 in this figure.



Example Figure 2. Dotplot of parameter estimates from a Bayesian model of bat viral shedding. This model, using data on adult female bats captured by EHA, was fit using Stan with a total of 4000 iterations. Means (dots) and 95% highest posterior density intervals (HPDI; black lines) for parameters are shown. Note the vertical dotted line representing a value of zero. The overall model intercept is extremely negative, indicating a low baseline viral detection rate. Interestingly, this data subset suggests that pregnant female bats are less likely to host viruses (the 95% HPDI for the Pregnancy Effect is entirely negative). Heterogeneity in the data attributable to various other biological and technological factors are being accounted and controlled for. These include country of specimen collection, year of specimen collection, diagnostic laboratory conducting the testing, specimen type, bat host species from which the specimen was collected, and viral family tested for. Magnitude of the variation along each of these dimensions is reflected in the mean parameter estimate for the corresponding variation terms in the model. Finally, the strength of the seasonality effect varies across bat host species, allowing for a seasonal trend in viral detection if the data supports such a pattern.